

Distributed Binary Event Detection Under Data-Falsification and Energy-Bandwidth Limitation

Edmond Nurellari, Des McLernon, *Member, IEEE*, Mounir Ghogho, *Senior Member, IEEE*, and Sami Aldalameh

Abstract—We address the problem of centralized detection of a binary event in the presence of falsifiable sensor nodes (SNs) (i.e., controlled by an attacker) for a bandwidth-constrained *under-attack* spatially uncorrelated distributed wireless sensor network (WSN). The SNs send their quantized test statistics over orthogonal channels to the fusion center (FC), which linearly combines them to reach a final decision. First (considering that the FC and the attacker do not act strategically), we derive (i) the FC optimal weight combining; (ii) the optimal SN to FC transmit power, and (iii) the test statistic quantization bits that maximize the probability of detection (P_d). We also derive an expression for the attacker strategy that causes the maximum possible FC degradation. But in these expressions, both the optimum FC strategy and the attacker strategy require *a-priori* knowledge that cannot be obtained in practice. The performance analysis of sub-optimum FC strategies is then characterized, and based on the (compromised) SNs willingness to collaborate, we also derive analytically the sub-optimum attacker strategies. Then, considering that the FC and the attacker now act strategically, we re-cast the problem as a minimax game between the FC and the attacker and prove that the Nash Equilibrium (NE) exists. Finally, we find this NE numerically in the simulation results and this gives insight into the detection performance of the proposed strategies.

Index Terms—Distributed detection, distributed processing, falsified sensor nodes, wireless sensor networks (WSN).

I. INTRODUCTION

CENTRALIZED detection of a binary event is one of the most important applications of wireless sensor networks (WSNs) [1], [2]. Multiple low-cost sensor nodes (SNs) are often spatially deployed over a specific field to observe such binary events. The SNs process the observed data and report back to a fusion center (FC) that optimally combines to reach a global decision. Being geographically dispersed to cover large areas, the SNs are constrained in both bandwidth and power. Moreover, SNs are usually unattended and this makes them vulnerable to different types of attacks. The overall detection performance strongly depends on the reliability of these SNs in the network. While fusing the data received by the spatially deployed SNs allows the FC to make a reliable decision, it is possible that one or more SNs (compromised by an attacker) deliberately falsify their local observations

to degrade the overall FC detection performance. However, there are a number of different approaches as to how the test statistics received from each SN can be efficiently used in order to achieve a reliable FC decision. Before introducing our proposed strategies, we will first give a brief review of related work.

The framework of distributed detection under *attack-free* WSNs has been extensively studied in [3]–[13], to name but just a few references. While [3]–[7] consider centralized detection by assuming WSNs with unlimited bandwidth/resources, the latter assumption was relaxed in [8]–[13] by considering centralized detection over bandwidth-constrained/energy-constrained WSNs. But these approaches are vulnerable to some security attacks as some of the SNs reporting to the FC may be compromised. As a result, the FC is not robust against such attacks and its detection performance may be degraded.

However, security issues in centralized detection using WSNs remain an open issue, see [14]–[19] and references therein. While there are many types of security threats, in this paper we focus on a single type of attack, which is the test statistic falsification (TSF) attack part of the Byzantine attacks family originally proposed by [20] and later widely used in the context of distributed detection (e.g., [19], [21], [22]).

Reference [22] characterizes the power of the attack analytically and a closed-form expression for the worst “detection error” is provided. Also, the minimum fraction of the compromised SNs that makes the FC incapable is derived. Reference [23] presents a technique to identify such compromised SNs and then to exclude them from contributing to the FC fusion process. In [24], a probabilistic TSF attack is proposed and the theoretical performance evaluation (in terms of destructiveness and stealthiness) is obtained. The authors of [25], in the context of smart grids, propose heuristic centralized algorithms to derive various strategies (attacker versus defender (i.e., FC) dynamics). Then, a distributed algorithm is proposed that guarantees convergence to the centralized solution taken at the FC.

Detection in the presence of binary falsification¹ (Byzantine) attacks is considered in [26]. Here, a reputation-based scheme is proposed for identifying the compromised SNs by accumulating the deviations between each SN and the FC decision over a time window duration. The authors in [27] also consider binary Byzantine attacks, in the context of target localization, where the SNs transmit their binary decisions to the FC. These authors also propose two techniques to mitigate

E. Nurellari and D. McLernon are with the School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT, Leeds, U.K. (e-mail: elen@leeds.ac.uk; d.c.mclernon@leeds.ac.uk).

M. Ghogho is with the School of Electronic and Electrical Engineering, University of Leeds, LS2 9JT, Leeds, U.K., and also with the International University of Rabat, Rabat, Morocco (e-mail: m.ghogho@leeds.ac.uk).

S. Aldalameh is with the Al-Zaytoonah University of Jordan, Amman, Jordan (e-mail: sami.dalameh@gmail.com).

This work was partially supported by EPSRC grant EP/N03435X/1 “Shaking Tunnel Vision”.

¹The compromised SNs falsify their hard decisions instead of their actual test statistics prior to transmission to the FC.

the negative input of the compromised SNs on the FC decision. However, identifying and then excluding the contributions of the compromised SNs from the FC decision process may not be the best strategy. For instance, we might end up removing (from contributing towards the global decision) compromised SNs that hold useful information (e.g., those SNs with high local SNRs). Furthermore, performing detection by means of one-bit SNs report combining at the FC is also not optimum.

Now, the publication closest to the work presented in this paper is [19], where an *under – attack* WSN framework over unlimited bandwidth is considered (i.e., infinite channel capacity) and the detection performance is investigated. But as the SNs are battery operated devices (i.e., limited power) and the bandwidth is finite, the assumption of infinite capacity is unrealistic. Furthermore, practical WSN scenarios suffer from fading and attenuation. The authors of [19] also do not propose any technique to mitigate the degradation caused by these compromised SNs.

So, the work in this paper investigates the detection performance of the *under – attack* energy-constrained/bandwidth-constrained WSNs. The compromised SNs (controlled by the attacker), are assumed to know the true hypothesis² (e.g., [19], [22]) and they use this *a – priori* knowledge to construct the most effective strategy to make the FC’s decision unreliable. For the FC, we assume that it is not compromised and receives the test statistic from both types of SNs (i.e., compromised and honest). The transmission (SNs to FC) links are modeled as flat fading, additive white Gaussian noise (AWGN) channels. The assumption of flat fading is reasonable as most of the WSNs operate at both short distances and low bit-rate due to resource limitations.

A. Contributions & Organization

While previous publications (as outlined above) have also examined sensor networks in the presence of falsified SNs, this paper deals with more realistic scenarios that include limited bandwidth fading channels, quantization of test statistics, etc. So our main contributions are as follows:

(i) We develop an efficient FC linear weight combining framework for an *under – attack* WSN that operates over limited bandwidth fading channels. The probability of detection (P_d) and the probability of false alarm (P_{fa}) based on this framework are derived in a closed-form. To maximize P_d for a fixed P_{fa} and to further reduce the optimisation complexity, we adopt the modified deflection coefficient (MDC) as an alternative function to be optimized and provide an optimisation problem to be solved from both the FC’s and the attacker’s perspective. Based on this optimisation problem (from the FC’s perspective), we derive analytically the optimal weight combining, the optimal SN to FC transmit power and the number of quantization bits for each SN. Unfortunately, these expressions require *a – priori* knowledge about the attacker parameters which cannot be attained in practice. Then (from the attacker’s perspective), we derive analytically (for a fixed number of compromised SNs) the optimum attacker strategy

²This leads to a conservative assessment but allows analytical tractability of the security risk.

which also depends upon the FC weight combining and the SNs transmit power.

(ii) So, motivated by the above, we next analyze the problem under different attacking and defending scenarios and characterize analytically the performance of sub-optimum strategies (from both the FC’s and the attacker’s perspective) that do not require knowledge of the FC mechanism and the attacker parameters. Also, based on the willingness of collaboration among the SNs (from the attacker’s perspective), we distinguish between two setups: a) all the SNs (compromised and honest) share their data with their neighbors, and b) just the compromised SNs are willing to collaborate among themselves to improve their attack strength.

(iii) Finally, we re-cast the problem as a minimax game between the FC and attacker and show that the NE (Nash Equilibrium) exists. Having defined the game, we use numerical simulations to find this NE point, thus identifying the optimum behavior of both the FC and the attacker in a game-theoretical sense.

Now, the summary of the paper is as follows. In Section II we describe the system model and provide a data transmission scheme. Section III describes the optimisation problem from both the FC’s and the attacker’s perspective. In Section IV we present our proposed attacker and FC strategies and in Section V we re-cast the problem and analyze the equilibrium. Finally, in Section VI we present simulation results and in Section VII we give some conclusions.

II. SYSTEM MODEL AND DATA TRANSMISSION

Consider the problem of detecting the presence of an unknown but deterministic signal $s(n)$ by an *under – attack* WSN consisting of M geographically distributed SNs and a FC (see Fig. 1). The honest SNs are represented with a black color and the compromised SNs (i.e., the ones controlled by the attacker) with a red color. The attacker’s aim is to successfully manipulate the FC global decision making process while the FC would like to detect reliably (i.e., with very high probability). Each SN collects N samples of the observed signal and performs energy estimation. Consistent with the underlying hypotheses, the measured signal ($s_i(n)$) at the i^{th} SN will be further corrupted by AWGN $w_i(n) \sim \mathcal{N}(0, \sigma_i^2)$:

$$\mathcal{H}_0 : x_i(n) = w_i(n) \quad (1)$$

$$\mathcal{H}_1 : x_i(n) = s_i(n) + w_i(n). \quad (2)$$

The i^{th} SN evaluates:

$$T_i = \sum_{n=1}^N (x_i(n))^2, \quad i = 1, 2, \dots, M \quad (3)$$

which for large N can be approximated by a Gaussian distribution [28]. It is not difficult to derive appropriate statistics in (4), where $\xi_i = \sum_{n=1}^N s_i^2(n) / N\sigma_i^2$. While the honest SNs transmit the actual test statistic (i.e., the true energies) to the FC, the compromised SNs falsify them before transmitting to the FC. Next we introduce the attacker model.

$$\mathbb{E}\{T_i|\mathcal{H}_0\} = N\sigma_i^2, \quad \text{Var}\{T_i|\mathcal{H}_0\} = 2N\sigma_i^4, \quad \mathbb{E}\{T_i|\mathcal{H}_1\} = N\sigma_i^2(1 + \xi_i), \quad \text{Var}\{T_i|\mathcal{H}_1\} = 2N\sigma_i^4(1 + 2\xi_i). \quad (4)$$

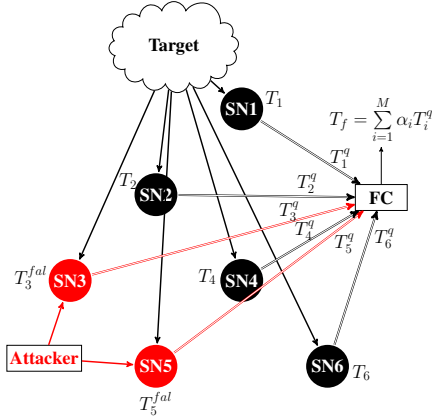


Fig. 1. Under attack schematic communication architecture between peripheral SNs and the fusion center (FC). Each SN generates a test statistic (T_i) by observing the target and can communicate with the FC only over an energy-constrained/bandwidth-constrained link. While the honest SNs (represented by black color) test statistics remain unchanged, the compromised SNs (represented by red color) falsify their test statistics to T_j^{fal} with $j = \{3, 5\}$ (where j is the compromised SN index) before transmitting to the FC.

A. Compromised SNs attack

In this work, the same attack model used in [19] is considered. The attacker (which has under its control a fraction (β) of the SNs) is assumed to know the true hypothesis² in (5) (e.g., [19], [22]). The remaining SNs are honest and completely unaware of the presence of falsified SNs. The i^{th} compromised SN falsifies its test statistic (T_i) before transmitting to the FC as follows:

$$T_i^{fal} = \begin{cases} T_i + C_i, & \text{under } \mathcal{H}_0 \\ T_i - C_i, & \text{under } \mathcal{H}_1 \end{cases} \quad (5)$$

where $C_i > 0$ is the parameter under the attacker's control. As we show later, there is an optimum C_i such that the detection performance back at the FC will be degraded the most. So, the test statistic (assuming compromised SNs) at the i^{th} SN can be represented as

$$\hat{T}_i = \begin{cases} T_i^{fal}, & \text{with probability } \beta \\ T_i, & \text{with probability } (1 - \beta) \end{cases} \quad (6)$$

where β is the fraction of the compromised SNs controlled by the attacker.

B. Data transmission

Now, because the SNs are battery operated devices (i.e., with limited on-board energy) then each SN i ($i = 1, 2, \dots, M$) has to quantize its test statistic (\hat{T}_i) to L_i bits prior to transmission to the FC. We assume that the FC is able to collect data from all the SNs via bandwidth constrained communication channels and furthermore, it is not itself compromised. As in [9], [13], we restrict the number of

quantization bits at the i^{th} SN to satisfy the channel capacity constraint:

$$L_i \leq \frac{1}{2} \log_2 \left(1 + \frac{p_i h_i^2}{\zeta_i} \right) \text{ bits} \quad (7)$$

where p_i denotes the transmit power of sensor i , h_i is the flat fading coefficient between SN i and the FC, and ζ_i is the variance of the AWGN at the FC. The quantized test statistic (T_i^q) at the i^{th} SN can be modeled (with L_i bits) as

$$T_i^q = \hat{T}_i + v_i \quad (8)$$

where v_i is quantization noise independent of $w_i(n)$ in (1) and (2). Assuming $T_i \in [0, 2U]$, then

$$\begin{cases} \hat{T}_i \in [C_i, 2U + C_i], & \text{under } \mathcal{H}_0 \text{ with probability } \beta \\ \hat{T}_i \in [-C_i, 2U - C_i], & \text{under } \mathcal{H}_1 \text{ with probability } \beta \\ \hat{T}_i \in [0, 2U], & \text{under } \{\mathcal{H}_p\}_{p=\{0,1\}} \text{ with probability } 1 - \beta. \end{cases} \quad (9)$$

Now, assuming a uniform quantizer with L_i bits (i.e., with a total of 2^{L_i} quantization levels), the quantizer step-size is always $\epsilon = \frac{2U}{2^{L_i}}$ and now v_i (see (8)) can be modeled as a r.v. uniformly distributed³ with $v_i \in [-\frac{\epsilon}{2}, \frac{\epsilon}{2}]$, where it is well-known that

$$\sigma_{v_i}^2 = \frac{U^2}{3 \times 2^{2L_i}}. \quad (10)$$

Note that the above analysis shows that the attacker (i.e., through the compromised SNs), does not introduce a larger quantization error noise (i.e., $\sigma_{v_i}^2$ in (10) remains the same as in the case of *attack-free* [9]). Now, linearly combining $\{T_i^q\}_{i=1}^M$ at the FC gives

$$T_f = \sum_{i=1}^M \alpha_i T_i^q \quad (11)$$

where the weights $\{\alpha_i\}_{i=1}^M$ will be optimized in Section III-A. For large M , the probability of detection (P_d) and the probability of false alarm (P_{fa}) can be approximated and

³This model that leads to (10) is only accurate for a relatively high number of bits (e.g., $L_i \geq 3$ in practice). For a smaller number of bits, the expression in (10) may not be very accurate but it is the only statistical measure available for such errors.

$$\mathbb{E}\{T_i^{fal}|\mathcal{H}_0\} = N\sigma_i^2 + C_i, \text{Var}\{T_i^{fal}|\mathcal{H}_0\} = 2N\sigma_i^4, \mathbb{E}\{T_i^{fal}|\mathcal{H}_1\} = N\sigma_i^2(1 + \xi_i) - C_i, \text{Var}\{T_i^{fal}|\mathcal{H}_1\} = 2N\sigma_i^4(1 + 2\xi_i). \quad (13)$$

shown to be respectively [19]:

$$\begin{aligned} P_d &= \Pr(T_f \geq \Lambda_f | \mathcal{H}_1) \\ &= \mathbf{1}^T \left(\mathbf{D}Q \left(\frac{\Lambda_f - \bar{\boldsymbol{\mu}}|\mathcal{H}_1}{\sqrt{\sum_{i=1}^M \alpha_i^2 (\text{Var}\{T_i|\mathcal{H}_1\} + \sigma_{v_i}^2)}} \right) \right) \\ P_{fa} &= \Pr(T_f \geq \Lambda_f | \mathcal{H}_0) \\ &= \mathbf{1}^T \left(\mathbf{D}Q \left(\frac{\Lambda_f - \bar{\boldsymbol{\mu}}|\mathcal{H}_0}{\sqrt{\sum_{i=1}^M \alpha_i^2 (\text{Var}\{T_i|\mathcal{H}_0\} + \sigma_{v_i}^2)}} \right) \right) \end{aligned} \quad (12)$$

with $\Lambda_f = \Lambda_f[1, 1, \dots, 1]_{2M}^T$ (Λ_f is the FC detection threshold); $\mathbf{1}_{2M}$ is a column vector of all ones; $\mathbf{D} = \text{diag}([\mathbf{b}_1 \odot \mathbf{b}_2 \odot \dots \odot \mathbf{b}_M])$ (\mathbf{b}_i is the i^{th} column vector of \mathbf{B} (where $\mathbf{B} = (1 - \beta)\mathbf{P} + \beta\mathbf{P}^c$) and \odot represents element-wise multiplication); the matrix \mathbf{P} is a binary matrix holding the 2^M possible combinations of M (compromised and honest) SNs on its rows with $(\mathbf{P})_{ij} = \{0, 1\}$ representing the compromised and honest SNs respectively (note that $(\mathbf{P})_{ij}$ represents the (i, j) element of \mathbf{P}); and \mathbf{P}^c is the element-wise (i.e., bitwise) logical complement of \mathbf{P} . Now, $\{\bar{\boldsymbol{\mu}}|\mathcal{H}_p\}_{p=\{0,1\}} = \mathbf{P}\{\boldsymbol{\mu}|\mathcal{H}_p\}_{p=\{0,1\}} + \mathbf{P}^c\{\boldsymbol{\mu}^{fal}|\mathcal{H}_p\}_{p=\{0,1\}}$ with $\{\boldsymbol{\mu}|\mathcal{H}_p\} = [\alpha_1 \mathbb{E}\{T_1|\mathcal{H}_p\}, \dots, \alpha_M \mathbb{E}\{T_M|\mathcal{H}_p\}]^T$ and $\{\boldsymbol{\mu}^{fal}|\mathcal{H}_p\} = [\alpha_1 \mathbb{E}\{T_1^{fal}|\mathcal{H}_p\}, \alpha_2 \mathbb{E}\{T_2^{fal}|\mathcal{H}_p\}, \dots, \alpha_M \mathbb{E}\{T_M^{fal}|\mathcal{H}_p\}]^T$ where $\mathbb{E}\{T_i|\mathcal{H}_p\}$ and $\mathbb{E}\{T_i^{fal}|\mathcal{H}_p\}$ are given in (4) and (13) respectively.

Finally, $Q(\cdot)$ represents the element-wise Q function operation. Next, we describe the optimisation problem under a power-constrained WSN.

III. FC AND ATTACKER PERFORMANCE OPTIMISATION UNDER A POWER-CONSTRAINED WSN

Now, if the attacker (which has under its control a fraction (β) of the M SNs) can successfully manipulate the FC global decision making process, the detection rate will be significantly low, the error rate in decision making will be high and the FC performance will be degraded. From the attacker's point of view, the more error it causes in the FC decision making, the more successful it is. The attacker has two available strategies: a) direct the compromised SNs to actually report their observation to the FC truthfully or b) direct the compromised SNs to falsify their observations prior to transmission to the FC. In the cases where the attacker decides to direct the compromised SNs to falsify their test statistics, what should be their optimum attacking parameter (C_i) ? We will answer this question in Section III-B.

From the FC's point of view, its data fusion mechanism should be robust and capable of defending against any attacking strategy adopted by any compromised SNs and directed by the attacker. The FC is aware that the attacker has an objective in conflict with its own (i.e., the FC tries to maximize the detection probability while the attacker tries to minimize it). However, the FC does not have any exact information about the attacking strategies. The only information available to the FC is: a) the quantized test statistics $\{T_i^q\}_{i=1}^M$ reported by M spatially distributed SNs, and b) the fraction⁴ (β) of these test statistics that are falsified. But it cannot recognize where these SNs are and estimate their "falsification parameter", C_i . So, the fusion data mechanism (based on this limited *a-priori* information) should be able to neutralize (or at least reduce) the impact of these compromised SNs.

So, in this Section, we would like to analyze the performance optimisation from the perspective of the FC and the attacker under a constraint of a maximum transmit power budget (P_t). Since the FC has under its control only the weight combiners $(\alpha_i, \forall i)$ in (11) and the SN to FC transmit power $(p_i, \forall i)$ in (7), its strategy is to maximize P_d with respect to the respective vectors containing these parameters (i.e., $\boldsymbol{\alpha}$ and \mathbf{p}). However, this is difficult and no closed-form solution can be obtained. Here, we introduce the MDC (which we will use later as an alternative function to be optimized). The MDC provides a good measure of the detection performance since it characterizes the variance-normalized distance between the centers of two conditional PDFs. This is given as:

$$\tilde{d}^2(\boldsymbol{\alpha}, \mathbf{p}) = \left(\frac{\mathbb{E}\{T_f|\mathcal{H}_1\} - \mathbb{E}\{T_f|\mathcal{H}_0\}}{\sqrt{\text{Var}\{T_f|\mathcal{H}_1\}}} \right)^2 = \frac{(\mathbf{b}^T \boldsymbol{\alpha})^2}{\boldsymbol{\alpha}^T \mathbf{R} \boldsymbol{\alpha}} \quad (14)$$

with the appropriate quantities given in (15) and (16) and where

$$\begin{aligned} \mathbf{b} &= [N\sigma_1^2\xi_1 - 2\beta C_1, \dots, N\sigma_M^2\xi_M - 2\beta C_M]^T \\ \boldsymbol{\alpha} &= [\alpha_1, \alpha_2, \dots, \alpha_M]^T, \mathbf{p} = [p_1, p_2, \dots, p_M]^T \\ \mathbf{R} &= \text{diag} \begin{bmatrix} 2N\sigma_1^4(1 + 2\xi_1) + \beta(1 - \beta)C_1^2 + \sigma_{v_1}^2 \\ 2N\sigma_2^4(1 + 2\xi_2) + \beta(1 - \beta)C_2^2 + \sigma_{v_2}^2 \\ \vdots \\ 2N\sigma_M^4(1 + 2\xi_M) + \beta(1 - \beta)C_M^2 + \sigma_{v_M}^2 \end{bmatrix}. \end{aligned} \quad (17)$$

A. FC performance optimisation

Now, the FC task (which knows that the WSN is *under-attack*) is to maximize the P_d (i.e., to detect with very high probability). We would like to make it clear that the FC knows⁴ β (i.e., knows the average percentage of compromised SNs

⁴In practice, the fraction representing the (on average) compromised SNs can be learned by observing the data sent by the SNs to the FC over a time window. But such an approach is beyond the scope of this work.

$$\mathbb{E}\{T_f|\mathcal{H}_0\} = \sum_{i=1}^M \alpha_i (N\sigma_i^2) + \sum_{i=1}^M \alpha_i (\beta C_i), \quad \mathbb{E}\{T_f|\mathcal{H}_1\} = \sum_{i=1}^M \alpha_i (N\sigma_i^2(1 + \xi_i)) - \sum_{i=1}^M \alpha_i (\beta C_i). \quad (15)$$

$$\text{Var}\{T_f|\mathcal{H}_0\} = \sum_{i=1}^M \alpha_i^2 (2N\sigma_i^4 + \beta(1 - \beta)C_i^2 + \sigma_{v_i}^2), \quad \text{Var}\{T_f|\mathcal{H}_1\} = \sum_{i=1}^M \alpha_i^2 (2N\sigma_i^4(1 + 2\xi_i) + \beta(1 - \beta)C_i^2 + \sigma_{v_i}^2). \quad (16)$$

(e.g., [19], [22])) but it cannot identify exactly who they are. Given the data fusion (11), the FC performs the following test:

$$\left. \begin{array}{l} \text{if } T_f < \Lambda_f, \text{ decide } H_0 \\ \text{if } T_f \geq \Lambda_f, \text{ decide } H_1 \end{array} \right\} \quad (18)$$

where Λ_f is the FC detection threshold. As we said earlier, the optimum weighting vector (α^o) and the optimum power allocation vector (p^o) that maximize P_d under the constraint of a maximum transmit power budget (P_t) are desired. More specifically (adopting the MDC), we require:

$$\begin{aligned} (\alpha^o, p^o) &= \arg \max_{\alpha, p} \left(\tilde{d}^2(\alpha, p) \right) \\ \text{subject to } &\sum_{i=1}^M p_i \leq P_t, \quad p_i \geq 0, \quad i = 1, 2, \dots, M. \end{aligned} \quad (19)$$

It is easily shown [9] that $\alpha^o = \mathbf{R}^{-1}\mathbf{b}$ with

$$\alpha_i^o = \frac{(\sigma_i^2 \xi_i - \frac{2\beta C_i}{N})}{2\sigma_i^4(1 + 2\xi_i) + \frac{\beta(1-\beta)C_i^2}{N} + \frac{\sigma_{v_i}^2}{N}}, \quad \forall i = 1, 2, \dots, M. \quad (20)$$

Note that the optimum weights $\{\alpha_i^o\}_{i=1}^M$ are a function of the SN transmit power (p_i) through the $\sigma_{v_i}^2$ terms (see (7) and (10)) and p_i is still to be optimized. We now substitute α^o back into (14) and solve the following optimisation problem

$$\begin{aligned} p^o &= \arg \max_p \left(\tilde{d}^2(\alpha^o, p) \right) \\ \text{subject to } &\sum_{i=1}^M p_i \leq P_t \text{ for } p_i \geq 0, \quad i = 1, 2, \dots, M. \end{aligned}$$

It can also be shown [9], that the above optimisation problem can be solved analytically by using the Lagrangian function and solving the appropriate K.K.T. conditions. The optimum SN to FC transmit power in this case (i.e., where the WSN is *under-attack*) can be shown to be

$$p_i^o = \left[\frac{U}{\sqrt{\lambda_0}} \sqrt{\frac{\zeta_i}{12h_i^2}} \left(\frac{\sigma_i^2 \xi_i - \frac{2\beta C_i}{N}}{\sigma_i^4(1 + 2\xi_i) + \beta(1 - \beta)\frac{C_i^2}{2N}} \right) - \frac{\frac{U^2 \zeta_i}{h_i^2}}{6N\sigma_i^4(1 + 2\xi_i) + 3\beta(1 - \beta)C_i^2} - \frac{\zeta_i}{h_i^2} \right]^+ \quad (21)$$

where $[y]^+$ equals 0 if $y < 0$, otherwise it equals y , and λ_0 is the Lagrangian multiplier that can be evaluated in a similar way as in [9] by imposing the equality constraint (i.e., $\sum_{i=1}^M p_i = P_t$) in (19). Now, (21) assumes that the FC knows the channel coefficients (h_i) for all SNs (honest and compromised). While the FC can obtain this information via a feedback from the honest SNs, the compromised SNs may

transmit to the FC wrong information regarding the channel. Nevertheless, here we assume that the compromised SNs only falsify their test statistics as in (5) and report true channel⁵ information to the FC. However, the channel information, for the compromised SNs, could be obtained by blind channel estimation techniques, etc., [29], [30]. Next, we analyze the performance optimisation from the attacker perspective.

B. Attacker performance optimisation

Now, the attacker would like to degrade as much as possible the FC detection performance. For a constant β (i.e., fraction of compromised SNs) the attacker plans the optimum C_i in (5) such that the FC becomes inefficient (i.e., useless). Adopting again the MDC (14), the optimisation problem can be expressed as:

$$C_i^o = \arg \min_{C_i} \left(\tilde{d}^2(\alpha_i, p_i, C_i) \right). \quad (22)$$

Note that (14) reaches its minimum value (i.e., zero) when $\mathbf{b}^T \alpha = \sum_{i=1}^M \alpha_i (N\sigma_i^2 \xi_i - 2\beta C_i) = 0$. Assuming that $C_i = C, \forall i$ (i.e., the same attack strength for all the compromised SNs) for simplicity, clearly the minimum of (14) can be achieved with

$$C^o = \sum_{i=1}^M \frac{\alpha_i N \sigma_i^2 \xi_i}{2\beta \sum_{i=1}^M \alpha_i}. \quad (23)$$

Now, this yields the maximum possible degradation that the attacker can cause to the FC. As can be seen, the optimum attacker strategy (C^o) is a function of the FC strategy (i.e., α_i in (11) which itself is a function of p_i through the $\sigma_{v_i}^2$ quantity (see (7), (10) and (20)). So, in order to adopt this strategy, the attacker needs to know α_i and $p_i, \forall i$. Since the FC is not compromised (i.e., still acts accordingly), these quantities cannot normally be obtained by the attacker.

As can be seen from the optimum FC weight protection strategy (20) and the attacker optimum strategy (23), there does not exist a dominant⁶ approach. Clearly the FC weights (α_i in (11)) depend on the attacker parameter C_i and vice versa. Next, we discuss in more detail the attacker versus the FC strategies and provide performance analysis in cases where limited *a-priori* knowledge about the attacker is available (i.e., without the need of exact knowledge of C_i).

⁵The channel estimation error (for both the honest and compromised SNs) can be modeled as a Gaussian random variable (i.e., $\hat{h}_{ij} = h_{ij} + e_h$) where $e_h \sim \mathcal{N}(0, \sigma_{e_h}^2)$ and \hat{h}_{ij} is the estimated flat fading channel coefficient.

⁶A dominant FC (attacker) strategy is said to be strictly dominant if it is the best strategy for the FC (attacker), no matter how the attacker (FC) decides to act.

IV. PERFORMANCE ANALYSIS

In this Section, starting with the optimum attacker strategy (23) and depending on the collaboration willingness among SNs and the available *a-priori* information that the attacker has about the FC combining strategy, we distinguish between two simulation setups in Section IV-A. Next, in Section IV-B we distinguish again between two different simulation setups but now from the perspective of the FC mechanisms.

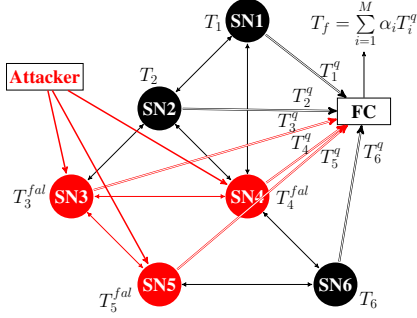


Fig. 2. Under attack schematic communication architecture among peripheral SNs and the FC. Similarly to Fig. 1, each SN generates a test statistic (T_i) by observing the target (not shown here for clearance purposes). While the honest SNs (black color) keep their test statistics unchanged, the compromised SNs (red color) directed by the attacker, will falsify their test statistics to T_j^{fal} with $j = \{3, 4, 5\}$ (where j is the compromised SN index). The SNs have partial connectivity among themselves (i.e., not a complete graph) (thin lines) and can communicate with the FC (thick lines) but only over an energy-constrained/bandwidth-constrained links.

A. Sub-optimum attacker's strategies

Here, we assume that the attacker knows that the FC uses a linear combining strategy but it is not aware of the combining weights used in (11). We also assume that the FC does not act strategically and uses weight combining, without trying to optimize against the behavior of compromised SNs. We now distinguish between the two following setups A-1 and A-2.

1) *Honest and compromised SNs collaborate (HCSC)*: Now, the optimum strategy (23) to be adopted by each compromised SN requires knowledge that cannot be obtained in practice as previously discussed. As a result, the attacker (not aware of α_i and p_i , $\forall i$) reasonably assumes equal combining at the FC (i.e., $\alpha_i = \frac{1}{M}$, $\forall i$) and directs the compromised SNs to attack with

$$C^{HCSC} = \frac{N}{M} \sum_{i=1}^M \frac{\sigma_i^2 \xi_i}{2\beta} \quad (24)$$

where the superscript “HCSC” refers to “*Honest and Compromised SNs Collaborate*”. However, the compromised SNs still require knowledge of σ_i^2 and ξ_i , $\forall i$ (to evaluate $\sum_{i=1}^M \sigma_i^2 \xi_i$) in order to implement the attacking strategy (24). When all the M SNs (honest and compromised) form a connected network⁷ and are willing to collaborate with each other (see Fig. 2), the quantity $\sum_{i=1}^M \sigma_i^2 \xi_i$ in (24) can be

⁷A connected network is any network where there is a path (i.e., over one or more links) between every pair of SNs in the network.

estimated using the average consensus algorithm [31]. Because of the communication topology for the M SNs (i.e., not fully connected), the average consensus algorithm ensures the availability of this term at each SN. The compromised SNs will still be camouflaged (i.e., unidentified) as they share with their neighbors just the true quantity $\sigma_i^2 \xi_i$ and the SNs cannot identify if their neighbors are honest or compromised.

2) *Compromised SNs (only) collaborate (CSC)*: Now, in the cases where not all of the M SNs (compromised and honest) are willing to collaborate with each other, the quantity $\sum_{i=1}^M \sigma_i^2 \xi_i$ in (23) cannot be obtained in practice. Note that the attacker has under its control just a fraction (β) ($\beta = \frac{F}{M} \leq 1$, where F is the number of falsified SNs of M SNs (see Fig. 2) and the other remaining honest SNs ($M - F$) do not share their observations with their neighbors. In this situation, the F compromised SNs collaborate with each other in order to estimate in a distributed fashion the $\sum_{i \in J} \sigma_i^2 \xi_i$ quantity, where J represents the compromised SNs set with cardinality F . Assuming that the F falsified SNs form a connected⁷ network, the average consensus algorithm [31] (like before) ensures the availability of this term at each falsified SN. After this stage, the compromised SNs attack (i.e., falsify their test statistics (3) as in (5)) with $C_i = C^{CSC}$, $\forall i$ with

$$C^{CSC} = \frac{N(M - F)}{M} \sum_{i \in J} \frac{\sigma_i^2 \xi_i}{2\beta} \quad (25)$$

where the superscript “CSC” refers to “*Compromised SNs (only) Collaborate*”.

B. Sub-optimum FC's strategies

Now, the optimum weights ($\alpha_i^o, \forall i$) in (20) are a function of the attacker parameter C_i which is difficult in practice (if not impossible) to obtain by the FC. In such a case, the FC adopts a sub-optimum but simple solution to minimize the degradation caused by the attacker. Assuming that the attacker does not act strategically (i.e., does not try to optimize against the FC approach) we now distinguish between the two following simulation setups B-1 and B-2.

1) *Weak attack FC based belief (WAFBB)*: Now, when the number of observed samples (N) is large and the FC believes that the attacker is directing the i^{th} compromised SN to attack with relatively small C_i , the FC weight combining can be approximated with

$$\alpha_i^{WAFBB} = \frac{\sigma_i^2 \xi_i}{2\sigma_i^4(1 + 2\xi_i) + \frac{\sigma_{v_i}^2}{N}}, \forall i = 1, 2, \dots, M \quad (26)$$

where the superscript “WAFBB” refers to “*Weak Attack FC Based Belief*” and the optimum SN to FC transmit power can be also approximated with

$$p_i^{WAFBB} = \left[\frac{U}{\sqrt{\lambda_0}} \sqrt{\frac{\zeta_i}{12h_i^2}} \left(\frac{\sigma_i^2 \xi_i}{\sigma_i^4(1 + 2\xi_i)} \right) - \frac{U^2 \zeta_i}{6N\sigma_i^4(1 + 2\xi_i)} - \frac{\zeta_i}{h_i^2} \right]^+ \quad (27)$$

Now, (26) and (27) coincide with the optimum weights and with the optimal SN transmit power allocation scheme respectively derived for the case of *attack-free* WSN in [9].

2) *Optimum attack FC based belief (OAFBB)*: Here, we consider the case when the FC believes that the attacker, with a fraction (β) of SNs under its control, attacks with the optimum parameter C^o (see (23)) (i.e., with $C_i = C^o$ in (5) but with $\alpha_i = \frac{1}{M}, \forall i$).

First of all, note that the FC knows that the compromised SNs (i.e., the attacker) have an alternative objective (i.e., the attacker would like to minimize, while the FC would like to maximize, the MDC in (14)) (i.e., the FC can work out the optimisation problem from the attacker perspective and evaluate (23)). Secondly, the FC concludes that the attacker cannot adopt this strategy in practice (since this optimum strategy requires $\alpha_i, \forall i$ and this parameter is controlled by the FC itself). In such a situation, it is reasonable that the FC believes that the attacker guides the compromised SNs to attack with C^o (see (23)) but with $\alpha_i = \frac{1}{M}, \forall i$. Now, the FC protection weights (α_i^{OAFBB}) can be shown to be (by substituting $C_i = \frac{N}{2\beta M} \sum_{i=1}^M \sigma_i^2 \xi_i$ in (20) and rearranging the terms):

$$\alpha_i^{OAFBB} = \frac{\sigma_i^2 \xi_i - \frac{1}{M} \sum_{i=1}^M \sigma_i^2 \xi_i}{2\sigma_i^4(1+2\xi_i) + N(1-\beta) \left(\frac{1}{2\sqrt{\beta M}} \sum_{i=1}^M \sigma_i^2 \xi_i \right)^2 + \frac{\sigma_{v_i}^2}{N}}. \quad (28)$$

The SN to FC transmit power (p_i^{OAFBB}) can be obtained in a similar way (by substituting $C_i = \frac{N}{2\beta M} \sum_{i=1}^M \sigma_i^2 \xi_i$ into (21)). Lastly, the superscript “*OAFBB*” refers to “*Optimum Attack FC Based Belief*”.

V. EQUILIBRIUM ANALYSIS

In this section, we consider the case where both the attacker and the FC act strategically and formulate the problem as a minimax game between two players, i.e., the attacker and the FC. The attacker has under its control one parameter (i.e., $C_i \forall i \in J$, with J defined in Section IV-A2) while the FC has control of the weight combining vector (i.e., α). As before, assuming $C = C_i$ (i.e., the same attack strength for each compromised SN) for simplicity, we first of all prove the existence of the Nash Equilibrium (NE)⁸ by showing that there exists a unique saddle-point in the minimax game between the attacker and the FC. Then, we find the optimum solution numerically by maximizing the deflection coefficient with respect to the FC weight combining parameter and then by minimizing it with respect to the attacker parameter (i.e., w.r.t. C). Next, we present a theorem, by help of which in Section V-A and Section V-B we prove the existence of NE. **Theorem 1** (Nikaido, [34]). Let $\mathcal{K}(x, y)$ be a pay-off function defined on the product space of \mathcal{X} by \mathcal{Y} , where \mathcal{X} and \mathcal{Y} are convex compact sets and continuous in each variable for

any fixed value of the other. If $\mathcal{K}(x, y)$ is quasi-concave in x and quasi-convex in y , then:

$$\max_{x \in \mathcal{X}} \min_{y \in \mathcal{Y}} \mathcal{K}(x, y) = \min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \mathcal{K}(x, y). \quad (29)$$

Next, we present the behavior of the MDC w.r.t. attacker strength C .

A. Modified deflection coefficient behavior with respect to C

In the next Lemma we prove the quasi-convexity behavior of the MDC w.r.t. C .

Lemma 1 : For a given α and \mathbf{p} , \tilde{d}^2 in (14) is a quasi-convex function of C .

Proof: The MDC can be written as:

$$\tilde{d}^2 = \frac{(x - 2\beta C b)^2}{y + dC^2} \quad (30)$$

$$\text{where } x = \sum_{i=1}^M \alpha_i (N\sigma_i^2 \xi_i), b = \sum_{i=1}^M \alpha_i, d = \sum_{i=1}^M \alpha_i^2 (\beta(1-\beta)), \\ y = \sum_{i=1}^M \alpha_i^2 (2N\sigma_i^4(1+2\xi_i) + \sigma_{v_i}^2).$$

Now considering α as a constant, differentiate \tilde{d}^2 w.r.t. C and by further simplification, we obtain:

$$\frac{\partial \tilde{d}^2}{\partial C} = \frac{(2\beta b C - x)(4\beta b y + 2x d C)}{(y + dC^2)^2} = 0. \quad (31)$$

So solving the above yields two critical points:

$$C_1^* = \frac{x}{2\beta b}, \quad C_2^* = -\frac{2\beta b y}{x d}. \quad (32)$$

Now, for a feasible attacker strength (i.e., for $C > 0$), the critical point C_1^* is feasible if $x, b > 0$ or $x, b < 0$. So, we have the following:

$$\left. \begin{array}{l} \text{if } x, b > 0 \text{ and for } C > C_1^*, f'(C) > 0 \\ \text{if } x, b < 0 \text{ and for } C > C_1^*, f'(C) > 0 \\ \text{if } x, b > 0 \text{ and for } C < C_1^*, f'(C) < 0 \\ \text{if } x, b < 0 \text{ and for } C < C_1^*, f'(C) < 0 \end{array} \right\} \implies \\ C_1^* \text{ is a global minimum.} \quad (33)$$

We also conclude that the other critical point C_2^* is not even a feasible point (i.e., $C_2^* < 0$) for $x, b > 0$ and $x, b < 0$. Hence, there is only one value of $C = C_1^*$ at which $\tilde{d}^2 = 0$. As a result, C_1^* being the unique global minimum $\implies \tilde{d}^2$ is a quasi-convex function of C .

B. Modified deflection coefficient behavior with respect to α and \mathbf{p}

Now, in Lemma 2, we show the behavior of \tilde{d}^2 in (14) from the perspective of the FC.

Lemma 2 : For a given C and \mathbf{p} , \tilde{d}^2 is a concave function of α .

⁸A Nash equilibrium, is a set of strategies, one for each player, such that no player has the incentive to unilaterally change its action. Players are in equilibrium if a change in strategies by any one of them would lead that player to earn less than if it remained with its current strategy.

Proof: Consider (14), then the Hessian of \tilde{d}^2 w.r.t. α (i.e., $\mathbf{H}_{\tilde{d}^2}$) can easily shown to be:

$$\begin{aligned} \mathbf{H}_{\tilde{d}^2} = & 2 \frac{\mathbf{b}\mathbf{b}^T}{\alpha\mathbf{R}\alpha} - 4 \frac{\mathbf{b}^T\alpha}{(\alpha\mathbf{R}\alpha)^2} (\mathbf{b}\alpha^T\mathbf{R} + \mathbf{R}\alpha\mathbf{b}^T) \\ & + 8 \frac{(\alpha^T\mathbf{b})^2}{(\alpha\mathbf{R}\alpha)^3} (\mathbf{R}\alpha\alpha^T\mathbf{R}) - 2 \frac{(\alpha^T\mathbf{b})^2}{(\alpha\mathbf{R}\alpha)^2} (\mathbf{R}). \end{aligned} \quad (34)$$

Now, to prove that \tilde{d}^2 is a concave function of α , we need to show [33]: $\alpha^T\mathbf{H}_{\tilde{d}^2}\alpha \leq 0, \forall\alpha$. This is given in Section VIII. From (36), $\alpha^T\mathbf{H}_{\tilde{d}^2}\alpha = 0, \forall\alpha \implies \tilde{d}^2$ is a concave function of α where the $\alpha_i^o, \forall i$ in (20) is the optimum solution. This concludes the proof.

Similarly, treating C (i.e., the attacker strength) fixed and for a given α (i.e., the weight combiner vector) it can be easily shown that \tilde{d}^2 is also a concave function of \mathbf{p} and p_i^o in (21) is the optimum solution. The proof is straightforward and we omit it here due to lack of space.

Now, since any concave function is quasi-concave, then by Theorem 1, a unique saddle-point exists in the minimax game which is the NE. We numerically evaluate this NE in the simulation results section.

VI. SIMULATION RESULTS

In this Section, the performances of the proposed strategies are evaluated numerically and compared to the *attack-free* scheme [9]. A WSN with a total of $M = 12$ SNs is considered (where a fraction of these SNs are compromised by the attacker with the same attacking strength (i.e., $C_i = C, \forall i$) for simplicity). We let $\sigma_i^2 = 0.1$, such that $\xi_a = 10 \log_{10} \left(\frac{1}{M} \sum_{i=1}^M \xi_i \right) = -10.5$ dB with arbitrarily chosen $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_M(n)] = [0.022, 0.0011, 0.18, 0.02, 0.0143, 0.0011, 0.0024, 0.2, 0.06, 0.09, 0.0143, 0.15]$ unless otherwise stated. The corresponding SN to FC channel gains are assumed to be ideally estimated (i.e., $\sigma_{e_h}^2 = 0$) for simplicity and are shown in Fig. 3. In addition we let $\zeta_i = 0.1, \forall i$. Finally, we choose L_i with equality in (7).

A. SN to FC optimal transmit power allocation and FC weight combining strategy

Now, we investigate the SN to FC transmit power for the optimum allocation scheme⁹ and the FC optimal weight combining strategy derived in Section III-A.

Fig. 3 (the middle plot) shows the optimal SN transmit power p_i^o for the i^{th} SN to the FC channel versus the attacker strength C and the lower plot shows the corresponding quantization bits. The actual channel coefficients (randomly chosen) are in the upper plot. Clearly, for the case of $C = 0$ (i.e., the *attack-free* scheme in [9]), more power is allocated to the SNs (i.e., SN3, SN8, SN9, SN10, and SN12) having both the

⁹The optimum SN power allocation scheme requires knowledge of the attacker strength C_i (see (21)). This is a strong assumption in practice and the exact knowledge of C_i cannot be attained in general. Nevertheless, here we consider this situation for performance comparison purposes and to create an idea about how the SN to FC transmit power allocation is affected.

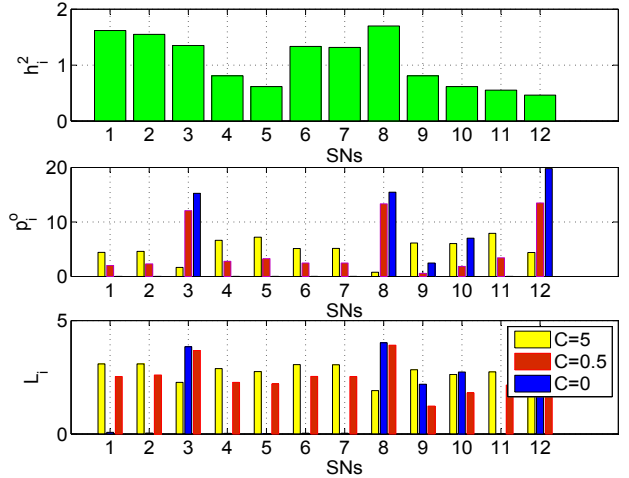


Fig. 3. SN optimal transmit power (p_i^o) and channel bit allocation (L_i) with $P_t = 60$, $U = 3$, $\xi_a = -10.5$ dB, $N = 20$, $\beta = 0.1$ and $\sigma_{e_h}^2 = 0$.

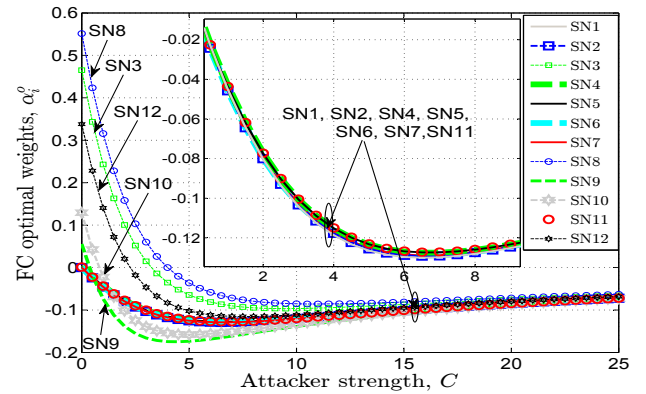


Fig. 4. FC optimal weights (α_i^o) versus the attacker strength (C) with $U = 3$, $\xi_a = -10.5$ dB, $P_t = 60$, $M = 12$, $N = 20$, $\beta = 0.1$ and $\sigma_{e_h}^2 = 0$.

best channels and high enough SNRs (ξ_i). Interestingly, those remaining SNs having very low SNRs (i.e., having useless local information) but having good (or bad) corresponding channels, are censored (i.e., do not transmit even a single bit). In this way, the SNs that have very bad channels (i.e., SNs that require very high power to transmit) or the SNs that have low SNRs (i.e., SNs that do not contain useful information) will be censored (i.e., will not transmit even one bit). This is not the case when $C = 0.5$ or $C = 5$ (we give an explanation later).

In Fig. 4 we investigate the FC combining response (with weight in (20)) versus attacker strength C . Clearly, when $C = 0$, the weights for the SNs permitted to transmit to the FC (i.e., SN3, SN8, SN9, SN10, and SN12) are greater than 0. As expected, the weights for the other remaining SNs are set to 0 (as these SNs are censored). Now, when C starts to increase, the FC response is decreasing the weights for all the SNs up to around $C = 5$ and allowing all the SNs to transmit to the FC (see Fig. 3 (middle plot)). However, for around $C > 5$, the FC response is by first increasing the weights for the SNs having low SNRs and as C gets larger, the FC combining strategy tends towards equal combining.

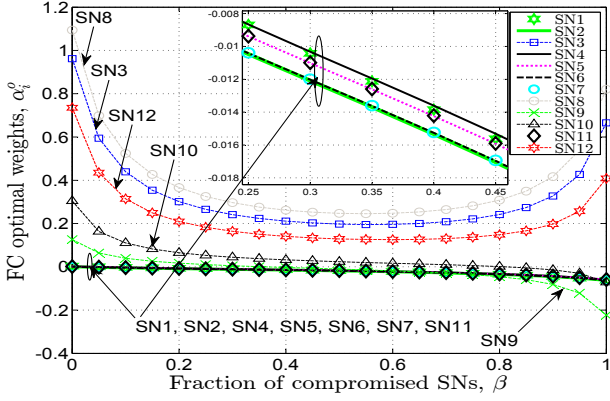


Fig. 5. FC optimal weights (α_i^o) versus fraction of the compromised SNs (β) with $U = 3$, $P_t = 60$, $N = 20$, $C_i = 0.1, \forall i$ and $\sigma_{e_h}^2 = 0$.

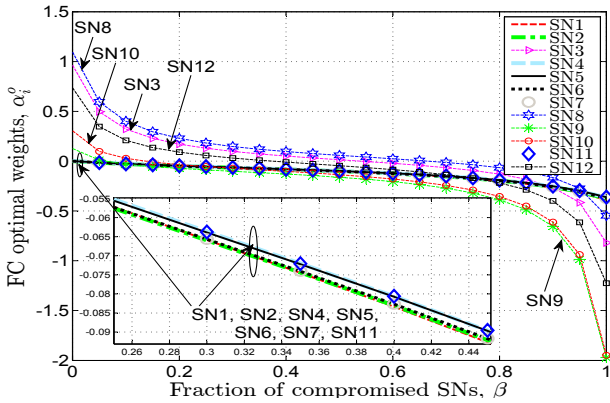


Fig. 6. FC optimal weights (α_i^o) versus fraction of the compromised SNs (β) with $U = 3$, $P_t = 60$, $N = 20$, $C_i = 0.6, \forall i$ and $\sigma_{e_h}^2 = 0$.

Similar to Fig. 4, in Fig. 5 (for $C = 0.1$) and in Fig. 6 (for $C = 0.6$) we plot the FC combining response (with weights in (20)) but now versus the fraction of compromised SNs (β). Interestingly, the optimal FC weight response for the *less informative* SNs (i.e., SN1, SN2, SN4, SN5, SN6, SN7, and SN11 classified by the power allocation scheme in the case of *attack-free* (i.e., $C = 0$)) remains almost constant with respect to β both in Fig. 5 and Fig. 6. However, that is not the case for the *more informative* SNs (i.e., SN3, SN8, SN9, SN10, and SN12). In Fig. 5, we observe that for the SNs 3, 8, and 12 (corresponding to the best SNRs) this relationship is convex while for the SNs 9 and 10 it is monotonically decreasing. Interestingly, in Fig. 6 (for a larger C) this relationship becomes monotonically decreasing for all the *more informative* SNs mentioned above.

The results provided in this Section cannot be attained in practice as the exact knowledge of C is required. However, they provide an insight as to how the FC power allocation and the weight combining strategy is influenced by both the attacker strength (C) and the compromised SNs fraction (β).

B. Detection performance for the proposed strategies

1) *Detection performance for fixed β* : Now, we investigate the detection performance of the proposed strategies described in Section IV-B for a fixed β .

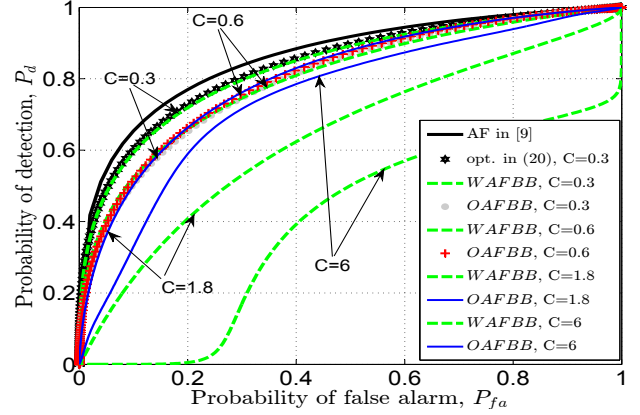


Fig. 7. Probability of detection (P_d) versus probability of false alarm (P_{fa}), with $U = 3$, $P_t = 60$, $M = 12$, $N = 20$, $\beta = 0.1$ and $\sigma_{e_h}^2 = 0$.

In Fig. 7, we show the receiver operating characteristic (ROC) parametrized on the attacker strength (C) for the proposed *WAFBB* and *OAFBB* strategies compared to the *attack free* (*AF*) case [9] (i.e., when there is no attack). We can observe that for $C = 0.3$ (as expected), the *WAFBB* strategy performs similar to the optimum strategy in (20) and better than *OAFBB* (up to $C = 0.6$) whereas after that, the *OAFBB* strategy dramatically outperforms the latter. We also note that for relatively very large C , it is possible to totally *blind* the FC when the *WAFBB* is used (i.e., to make it incapable of detecting) but only when the WSN operates at low probability of false alarm (P_{fa}).

Now, we would like to emphasize that the *WAFBB* strategy has particular importance when the FC does not have any *a-priori* knowledge about the β and C parameters. But the *OAFBB* strategy requires just knowledge of the compromised SNs fraction⁴ (β) which is possible to be obtained by the FC in practice.

2) *Detection performance for fixed C* : Now, we investigate the detection performance of the proposed strategies described in Section IV-B for a fixed C .

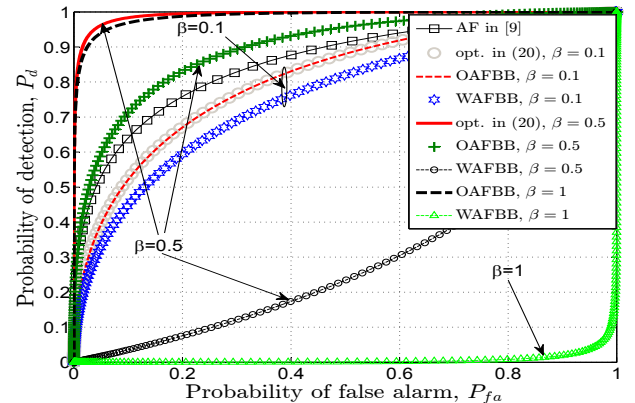


Fig. 8. Probability of detection (P_d) versus probability of false alarm (P_{fa}) with $U = 3$, $P_t = 60$, $M = 12$, $N = 20$, $C_i = 0.9, \forall i$ and $\sigma_{e_h}^2 = 0$.

In Fig. 8, we show the ROC performance for the two different proposed strategies (parametrized on β) compared to the optimum strategy in (20) and *AF* in [9]. We can observe

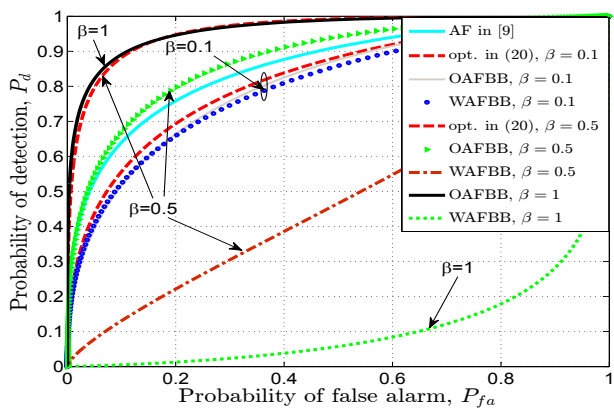


Fig. 9. Probability of detection (P_d) versus probability of false alarm (P_{fa}) with $U = 3$, $P_t = 60$, $M = 12$, $N = 20$ and $C_i = 0.6, \forall i$, and $\sigma_{e_h}^2 = 0$.

that for small β (more specifically $\beta = 0.1$), both the optimum and *OAFBB* strategies outperform the *WAFBB* strategy and their performances are worst than the *AF* performance. Interesting, when β increases (more specifically, $\beta = 0.5$ and $\beta = 1$), both the optimum and *OAFBB* strategies outperform the *AF* detection performance for all the values of P_{fa} and their detection performances improve proportionally with β . This is as expected, since from the attacker's perspective there does exist an optimum β that most degrades the FC's detection capability (Fig. 10 later on captures and demonstrates this behavior better) (see also (23)). We would also like to make it clear that the FC's ROC performance behavior depends not only on the compromised SNs fraction (β) but also on the attacker strength (C_i). Furthermore, the optimum β that causes the maximum FC's detection degradation depends itself in C_i (see (23)). Deviating from this optimum strategy (i.e., (β, C_i)), the attacker might help the FC to further utilize its detection performance rather than causing degradation. While from $\beta = 0.5$ to $\beta = 1$ the performance of the optimum strategy in (20) and the *OAFBB* strategy improves, that is not the case when *WAFBB* is used (its performance degrades) and when $\beta = 1$ it is sufficient to *blind* the FC even when the WSN operates at a relatively high P_{fa} . Now, this is as expected because the *WAFBB* requires no *a-priori* knowledge regarding the attacker's parameters.

In Fig. 9, we investigate the same situation as for Fig. 8 but now for $C = 0.6$. In this case (when $\beta = 0.1$), the optimum strategy slightly outperforms the *OAFBB* and *WAFBB* strategies. However, similar to Fig. 8, when more than 50% of SNs are compromised, the *OAFBB* strategy significantly outperforms the *WAFBB* strategy. Furthermore, its detection performance improves (for $\beta \geq 0.5$) proportionally as β increases.

In Fig. 10, we again show the ROC as a function of β but now for a lower C (more specifically for $C = 0.2$). As expected, the *WAFBB* performs similar to the optimum strategy and outperforms the *OAFBB* at low β and C , as the *WAFBB* is derived under these assumptions. Also, we can observe that the ROC performance of the optimum and *OAFBB* strategies (when $\beta = 0.1$ is used) outperforms those when $\beta = 0.5$. This is an intuitive result as the smaller the

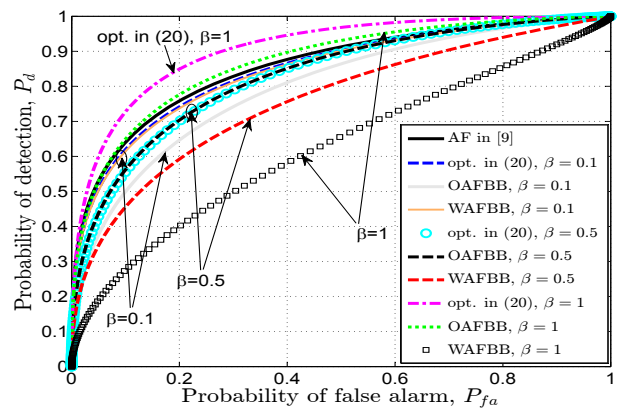


Fig. 10. Probability of detection (P_d) versus probability of false alarm (P_{fa}) with $U = 3$, $P_t = 60$, $M = 12$, $N = 20$, $C_i = 0.2, \forall i$ and $\sigma_{e_h}^2 = 0$.

fraction of compromised SNs participating in the network, the better is the FC's detection performance. However, for $\beta \geq 0.5$, the detection performance is shown to be improving with β whereas for the *WAFBB* strategy its performance degrades as β increases. Interestingly, when 50% of the SNs are compromised, both the optimum and *OAFBB* strategies perform in a similar manner.

It is now clear that (from the attacker perspective) there is an optimum number of compromised SNs (fraction β) that causes the maximum FC's detection performance degradation when using the *OAFBB* and the optimum FC strategy in (20). On the contrary, when the *WAFBB* strategy is considered, we conclude that the FC's detection performance degrades as β increases. However, it has particular importance in practice as no *a-priori* knowledge for the attacker parameters is required.

C. Equilibrium analysis of minimax game

In this section we analyze the equilibrium point of the minimax game and find the Nash Equilibrium (NE). The NE is the maximum probability of detection considering the FC's best linear weight combining strategy (joint optimization of α, \mathbf{p}) against attacker's strategy (i.e., C for a given fraction of compromised SNs β).

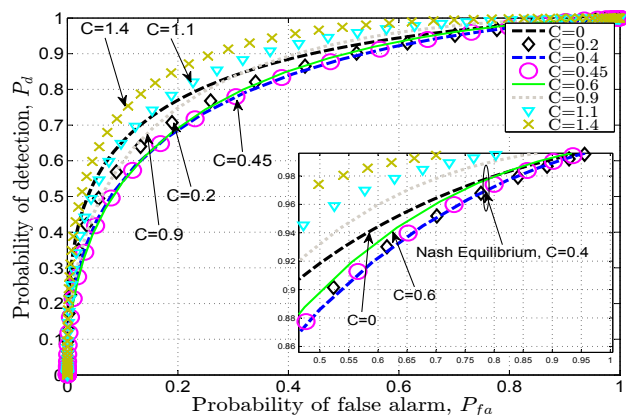


Fig. 11. Probability of detection (P_d) versus probability of false alarm (P_{fa}), with $U = 3$, $\xi_a = -10.5$ dB, $P_t = 60$, $M = 12$, $N = 20$, $\beta = 0.2$, $\sigma_{e_h}^2 = 0$ and with optimum weights in (20).

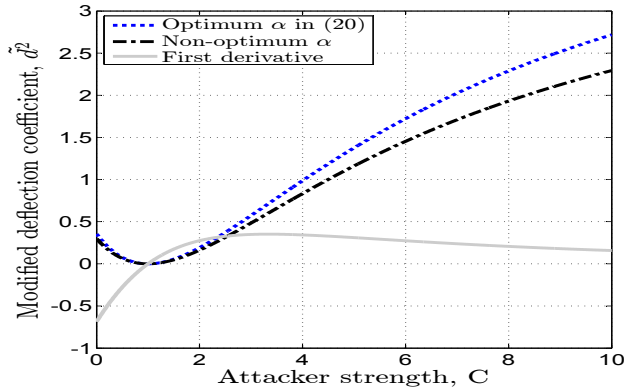


Fig. 12. Modified deflection coefficient (\bar{d}^2) versus the attacker strength (C) with $U = 3$, $\xi_\alpha = -10$ dB, $s_i = 0.1, \forall i$, $P_t = 60$, $M = 12$, $N = 20$, $\beta = 0.1$ and $\sigma_{eh}^2 = 0$.

In Fig. 11 the ROC behavior against the attacker's strength and the FC's combining weights is shown. As expected (see Section V on equilibrium analysis) there does exist a NE and it is shown to occur for the pair $C = 0.4$ and α^o (with α_i^o in (20)). Clearly, from the attacker's perspective, this strategy causes the maximum detection performance degradation $\forall P_{fa}$ and deviating from this strategy will not benefit the attacker.

Now, in Fig. 12, the modified deflection coefficient (MDC) against the attacker strength (C) is shown for two examples (i.e., with the optimum FC weights combining in (20) and the non-optimum weight combining drawn from the uniform distribution (i.e., $\alpha_i \sim \mathcal{U}(0, 1)$ in (11)). We can observe that the NE is shown to occur at $C = 1$ and deviating from this point (i.e., this strategy) the attacker will not benefit (i.e., it will not gain in terms of the FC's performance degradation). It is also clear that if the FC deviates from the optimum combining strategy (i.e., from the weights α_i^o in (20)), its detection performance will be worst or at least will not improve $\forall C$.

VII. CONCLUSION

In this paper, we have addressed the problem of distributed detection by an *under-attack* WSN that operates over limited bandwidth communication fading channels. Based on a simple linear weight combining rule at the FC and adopting the modified deflection coefficient (as an alternative function to be optimized), we give closed-form expressions for the optimal FC combining weights, the SN to FC transmit power allocation, and the test statistics quantization bits. The attacker optimal strategy is also derived and shown to be dependent on the FC combining weights. Furthermore, sub-optimum FC strategies (based on weight combining and the SN transmit power) that do not require the exact knowledge of the attacker strength C are also derived and analyzed.

We have also analyzed the equilibrium to the minimax problem and have proved that the Nash Equilibrium (NE) exists and found this optimal solution numerically in the simulation results. We compare our proposed FC strategies with the one derived under an *attack-free* scenario and show significant detection performance improvement.

Future work will consider a general (non-linear) optimal combining strategy at the FC and study attackers that (unlike

in this paper) do not know the true state of the target (i.e., they are less dangerous attackers).

VIII. PROOF OF $\alpha^T \mathbf{H}_{\bar{d}^2} \alpha \leq 0, \forall \alpha$ IN (34)

Consider the $\mathbf{H}_{\bar{d}^2}$ in (34), and show that $\alpha^T \mathbf{H}_{\bar{d}^2} \alpha = 0, \forall \alpha$. *Proof:* Multiplying (34) from the left by α^T and from the right by α , we get:

$$\begin{aligned} \alpha^T \mathbf{H}_{\bar{d}^2} \alpha &= 2 \frac{\alpha^T \mathbf{b} \mathbf{b}^T \alpha}{\alpha^T \mathbf{R} \alpha} - 4 \frac{\alpha^T \mathbf{b}^T \alpha}{(\alpha^T \mathbf{R} \alpha)^2} (\mathbf{b} \alpha^T \mathbf{R} + \mathbf{R} \alpha \mathbf{b}^T) \alpha \\ &+ 8 \frac{\alpha^T (\alpha^T \mathbf{b})^2}{(\alpha^T \mathbf{R} \alpha)^3} (\mathbf{R} \alpha \alpha^T \mathbf{R}) \alpha - 2 \frac{(\alpha^T \alpha^T \mathbf{b})^2}{(\alpha^T \mathbf{R} \alpha)^2} (\mathbf{R}) \alpha. \end{aligned} \quad (35)$$

Rearranging the terms and by further simplification, we obtain:

$$\begin{aligned} \alpha^T \mathbf{H}_{\bar{d}^2} \alpha &= 2 \frac{\alpha^T \mathbf{b} \mathbf{b}^T \alpha}{\alpha^T \mathbf{R} \alpha} - 8 \frac{\mathbf{b}^T \alpha \alpha^T \mathbf{b}}{\alpha^T \mathbf{R} \alpha} \\ &+ 8 \frac{\mathbf{b}^T \alpha \alpha^T \mathbf{b}}{\alpha^T \mathbf{R} \alpha} - 2 \frac{\alpha^T \mathbf{b} \mathbf{b}^T \alpha}{\alpha^T \mathbf{R} \alpha} = 0. \end{aligned} \quad (36)$$

This concludes the proof.

REFERENCES

- [1] D. Estrin, L. Girod, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," *Proc. ICASSP*, Salt Lake City, UT, United States, May 2001.
- [2] O. Songhwa, C. Phoebus, M. Michael, M. Srivastava, and S. Shankar, "Instrumenting Wireless Sensor Networks for Real-time Surveillance," *Proc. of the Int. Conf., on Robotics and Automation (ICRA)*, May 2006.
- [3] P. K. Varshney, *Distributed Detection and Data Fusion*, 1st edn., Springer, New York, 1997.
- [4] J. N. Tsitsiklis, "Decentralized detection," in *Advances in Signal Process.*, H. V. Poor and J. B. Thomas, Eds. New York: JAI, 1993, vol. 2, pp. 297-344.
- [5] R. Blum, S. Kassam, and H. Poor, "Distributed detection with multiple sensors: Part II-advanced topics," *Proc. IEEE*, vol. 85, no. 1, pp. 64-79, Jan. 1997.
- [6] Z. Quan, S. Cui, and A. H. Sayed, "Optimal linear cooperation for spectrum sensing in cognitive radio networks," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 28-40, Feb. 2008.
- [7] S. Barbarossa, S. Sardellitti, and P. Di Lorenzo, "Distributed Detection and Estimation in Wireless Sensor Networks," In Rama Chellappa and Sergios Theodoridis eds., Academic Press Library in Signal Process., vol. 2, Communications and Radar Signal Process., pp. 329-408, 2014.
- [8] J. F. Chamberland and V. V. Veeravalli, "Asymptotic results for decentralized detection in power constrained wireless sensor networks," *IEEE J. Sel. Areas in Comm.*, vol. 22, no. 6, pp. 1007-1015, Aug. 2004.
- [9] E. Nurellari, D. McLernon, M. Ghogho, and S. Aldalameh, "Optimal quantization and power allocation for energy-based distributed sensor detection," *Proc. EUSIPCO*, Lisbon, Portugal, 1-5 Sept. 2014.
- [10] A. Ribeiro and G. B. Giannakis, "Bandwidth-constrained distributed estimation for wireless sensor networks, part I: Gaussian case," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp.1131-1143, 2006.
- [11] E. Nurellari, D. McLernon, and M. Ghogho, "Distributed Two-Step Quantized Fusion Rules via Consensus Algorithm for Distributed Detection in Wireless Sensor Networks," in *IEEE Trans., Signal and Information Process., over Networks*, vol. PP, no. 99, pp. 1-1, Apr. 2016.
- [12] X. Zhang, H. V. Poor, and M. Chiang, "Optimal power allocation for distributed detection over MIMO channels in wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4124-4140, Sep. 2008.
- [13] E. Nurellari, S. Aldalameh, M. Ghogho and D. McLernon, "Quantized Fusion Rules for Energy-Based Distributed Detection in Wireless Sensor Networks," *Proc. SSPD*, Edinburgh, Scotland, 8-9 Sept. 2014.
- [14] L. Zhang, G. Ding, Q. Wu, Y. Zou, Z. Han, and J. Wang, "Byzantine Attack and Defense in Cognitive Radio Networks: A Survey," *IEEE Communications Surveys & Tutorials*, vol. PP, issue: 99, Apr. 2015.
- [15] F. Yu, H. Tang, M. Huang, and Z. Li, "Defense against spectrum sensing data falsification attacks in mobile ad hoc networks with cognitive radios," *Proc. MILCOM*, Boston, MA, 18-21 Oct. 2009.

- [16] T. Zhao and Y. Zhao, "A new cooperative detection technique with malicious user suppression," *Proc. ICC*, Germany, 14-18 Jun. 2009.
- [17] A. Vempaty, L. Tong, and P. Varshney, "Distributed Inference with Byzantine Data: State-of-the-Art Review on Data Falsification Attacks," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 65-75, 2013.
- [18] Y. Cai, Y. Mo, K. Ota, C. Luo, M. Dong, and L. Yang, "Optimal data fusion of collaborative spectrum sensing under attack in cognitive radio networks," *IEEE Network*, vol. 28, no. 1, pp. 17-23, Jan-Feb. 2014.
- [19] B. Kaikhura, S. Brahma, and P. K. Varshney, "On the performance analysis of data fusion schemes with Byzantines," *Proc. ICASSP*, May 2014.
- [20] L. Lamport, R. Shostak, and M. Pease, "The byzantine generals problem," *ACM Trans. Program. Lang. Syst.*, vol. 4, no. 3, pp. 382-401, Jul. 1982. [Online]. Available: <http://doi.acm.org/10.1145/357172.357176>.
- [21] V. S. S. Nadendla, Y. S. Han, and P. K. Varshney, "Distributed Inference With M-Ary Quantized Data in the Presence of Byzantine Attacks," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2681-2695, May 2014.
- [22] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of byzantine attacks," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 16-29, Oct. 2009.
- [23] P. Kaligineedi, M. Khabbazian, and V. K. Bhargava, "Secure cooperative sensing techniques for cognitive radio systems," *Proc. ICC*, pp. 3406-3410, Beijing, China, 19-23 May 2008.
- [24] L. Zhang, Q. Wu, G. Ding, S. Feng, and J. Wang, "Performance analysis of probabilistic soft SSDF attack in cooperative spectrum sensing," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, pp. 81, May 2014.
- [25] S. Cui, Z. Han, S. Kar, T. T. Kim, H. Poor, and A. Tajer, "Coordinated data-injection attack and detection in smart grid," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 106-115, Sept. 2012.
- [26] A. S. Rawat, P. Anand, H. Chen, and P. K. Varshney, "Collaborative spectrum sensing in the presence of byzantine attacks in cognitive radio networks," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 774-786, Jan. 2011.
- [27] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. Varshney, "Localization in wireless sensor networks: Byzantines and mitigation techniques," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1495-1508, Mar. 2013.
- [28] H. Urick, "Energy detection of unknown deterministic signals," *Proc. IEEE*, vol. 55, pp. 523-531, Apr. 1967.
- [29] A. Taherpour, M. N. Kenari, and S. Gazor, "Multiple antenna spectrum sensing in cognitive radios," *IEEE Trans. Wireless Commun.*, vol. 9, no. 2, pp. 814-823, Feb. 2010.
- [30] Z. Chen, T. Cooklev, C. Chen, and C. Pomalaza-Raez, "Modeling primary user emulation attacks and defenses in cognitive radio networks," *Proc. IEEE International Performance Comput. Commun. Conf.*, pp. 208-215, Dec. 2009.
- [31] R. O. Saber, J. A. Fax, R. M. Murray, "Consensus and cooperation in networked multi-agent systems," *Proc. IEEE*, 95(1), pp. 215-233, Jan. 2007.
- [32] S. M. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory*, Englewood Cliffs, NJ: Prentice-Hall PTR, 1993.
- [33] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge Univ. Press, 2003.
- [34] H. Nikaido and L. Vandenberghe, "On von neumann's minimax theorem," *Pacific Journal of Mathematics*, vol. 4, no. 1, pp. 65-72, 1954.

PLACE
PHOTO
HERE

Edmond Nurellari received the Diploma and the M.Sc. degree in Electrical and Electronic Engineering both from Eastern Mediterranean University, Northern Cyprus, in 2010 and in 2012 respectively.

From September 2010 to February 2013, he served as a Research and Teaching Assistant in the department of Electrical and Electronic Engineering at Eastern Mediterranean University. In 2013, he was awarded the Leeds International Research Scholarship (LIRS) to pursue his Ph. D. at the School of Electronics and Electrical Engineering, University of

Leeds, U.K.

His research interests includes distributed signal processing, signal processing on graphs, resource allocations and distributed decisions in wireless sensor networks. He has served as an Invited Reviewer for the IEEE Transactions on Signal and Information Processing over Networks, IEEE Communication Letter, Springer's Wireless Networks Journal, and IEEE Flagship conferences.

PLACE
PHOTO
HERE

Des McLernon (M'94) received his B.Sc in Electronic and Electrical Engineering and his M.Sc. in Electronics, both from Queens University of Belfast, N. Ireland.

He then worked on radar research and development with Ferranti Ltd. in Edinburgh, Scotland, and later joined the Imperial College, University of London, where he received his Ph.D. in signal processing. After first lecturing at South Bank University, London, UK, he moved to the School of Electronic and Electrical Engineering at the University of Leeds, UK, where he is a Reader in Signal Processing and Director of Graduate Studies.

His research interests are broadly within the domain of signal processing for communications, in which area he has published over 285 journal and conference papers.

PLACE
PHOTO
HERE

Mounir Ghogho (SM'96) received the MSc degree (DEA) in 1993 and the PhD degree in 1997 from the National Polytechnic Institute of Toulouse, France. He was an EPSRC Research Fellow with the University of Strathclyde, Glasgow (Scotland), from September 1997 to November 2001.

Since December 2001, he has been a faculty member with the school of Electronic and Electrical Engineering at the University of Leeds (England), where he is currently a Professor. He is also currently a Research Director and a Scientific Advisor

to the President at the International University of Rabat (Morocco). He was awarded the UK Royal Academy of Engineering Research Fellowship in September 2000. He is a recipient of the 2013 IBM Faculty Award. He is currently an Associate Editor of the Signal Processing Magazine. He served as an Associate Editor of the IEEE Signal Processing Letters from 2001 to 2004, the IEEE Transactions on Signal Processing from 2005 to 2008, and the Elsevier Digital Signal Processing journal from 2011 to 2012. He served as a member of the IEEE Signal Processing Society SPCOM Technical Committee from 2005 to 2010, a member of IEEE Signal Processing Society SPTM Technical Committee from 2006 to 2011, and is currently a member of the IEEE Signal Processing Society SAM Technical Committee. He was the General Chair of the European Signal Processing conference Eusipco2013 and the IEEE workshop on Signal Processing for Advanced Wireless Communications SPAWC2010, the technical co-chair of the MIMO symposium of IWCMC 2007 and IWCMC 2008, and a technical area co-chair of Eusipco 2008, Eusipco 2009 and ISCCSP05. He is the general Chair of IEEE WCNC 2019.

His research interests are in signal processing and communication networks. He has published over 260 journal and conference papers. He held invited scientist/professor positions at Telecom Paris-Tech (France), NII (Japan), BUPT (China), University Carlos 3rd of Madrid (Spain), ENSICA (Toulouse), Darmstadt Technical University (Germany), and Minnesota University (USA). He is the Eurasip Liaison in Morocco.

PLACE
PHOTO
HERE

Sami Aldalameh is an assistant professor in the department of communication and computer engineering at Al-Zaytoonah University of Jordan since 2013.

He received his PhD degree from University of Leeds, UK, in 2013. He was awarded his MSc degree in Communications Signal Processing from King's College London, UK, in 2007 with distinction and top of his class. He earned his BSc degree in Electrical Engineering from the University of Jordan, Jordan, in 2006.

His research interests include distributed detection and estimation, wireless sensor networks, stochastic geometry models for wireless networks, delay and throughput in wireless sensor networks, and sensor signal processing. Dr. Aldalameh is an active IEEE member for 13 years with affiliation in the IEEE signal processing, communication, and computer societies.